Course One
Foundations of Data Science

Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. You can use this document as a guide to consider your responses and reflections at different stages of the data analytical process. Additionally, the PACE strategy documents can be used as a resource when working on future projects.

Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

- ☐ Complete the PACE Strategy Document to plan your project while considering your audience members, teammates, key milestones, and overall project goal.
- ☐ Create a project proposal for the data team.

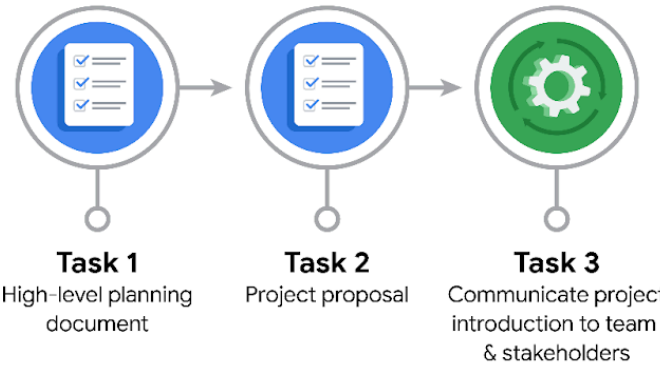
Relevant Interview Questions

Completing this end-of-course project will empower you to respond to the following interview topics:

- As a new member of a data analytics team, what steps could you take to get 'up to speed' with a current project? What steps would you take? Who would you like to meet with?
- How would you plan an analytics project?
- What steps would you take to translate a business question to an analytical solution?
- Why is actively managing data an important part of a data analytics team's responsibilities?
- What are some considerations you might need to be mindful of when reporting results?

Reference Guide

This project has three tasks; the following visual identifies how the stages of PACE are incorporated across those tasks.



Data Project Questions & Considerations

PACE: Plan Stage

- Who is your audience for this project?

The audience for this project includes the TikTok data team, specifically data analysts, data scientists, and stakeholders involved in content moderation and machine learning. This team will be interested in the project's ability to enhance content moderation through effective claims classification.

- What are you trying to solve or accomplish? And, what do you anticipate the impact of this work will be on the larger needs of the client?

The project aims to develop a machine learning model to classify user interactions as claims or opinions. This will streamline content moderation by automating the process, thus reducing the burden on human moderators and improving the accuracy and efficiency of content review.

- What questions need to be asked or answered?

  - What are the specific criteria for classifying interactions as claims or opinions?
  - What data sources are available for training and testing the model?
  - What are the expected performance metrics for the model (e.g., accuracy, precision, recall)?
  - How will the model be integrated into the existing moderation workflow?

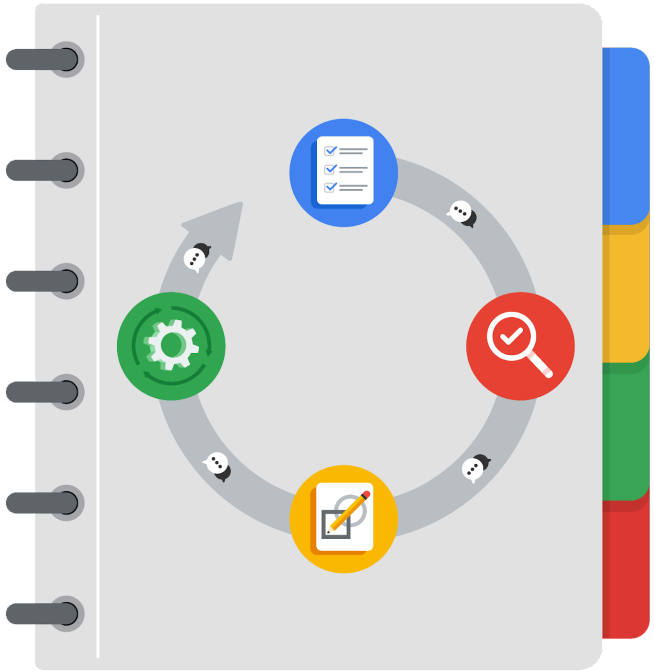
- What resources are required to complete this project?

  - Access to user interaction data on TikTok
  - Machine learning tools and libraries
  - Computational resources for model training and evaluation
  - Expertise in data science and machine learning
  - Collaboration with content moderation experts for domain knowledge

- What are the deliverables that will need to be created over the course of this project?

  - Data exploration and cleaning reports
  - A trained machine learning model for classification
  - Evaluation metrics and results
  - A final project proposal and documentation
  - Integration plan for the model with existing moderation systems

THE PACE WORKFLOW



[Alt-text: The PACE Workflow with the four stages in a circle: plan, analyze, construct, and execute.]

You have been asked to demonstrate for the company's data team how you would use the PACE workflow to organize and classify tasks for the upcoming project. Select a PACE stage from the dropdown buttons. A few tasks involve more than one stage of the PACE workflow. Additionally, not every workplace scenario will require every task. Refer back to the Course 1 end-of-course portfolio project overview reading if you need more information about the tasks within the project.

Project tasks

Following are a group of tasks your company's data team has determined need to be completed within this project. The data analysis manager has asked you to organize these tasks in preparation for the project proposal document. First, identify which stage of the PACE workflow each task would best fit under using the drop down menu. Next, give an explanation of why you selected the stage for each task. Review the following readings to help guide your selections and explanation: The PACE stages and Communicate objectives with a project proposal. You will later reorder these tasks within a project proposal.

1. **Evaluating the model:** Execute

   Why did you select this stage for this task?

Evaluating the model involves testing its performance and ensuring it meets the project's goals and criteria. This falls under the Execute stage as it requires applying the model and interpreting its results

2. **Conduct hypothesis testing:** Analyze **and** Construct

   Why did you select these stages for this task?

Hypothesis testing involves analyzing the data to validate assumptions and ensure the model's accuracy. This falls under the Analyze stage as it requires analyzing data patterns and model behavior.

3. **Begin exploring the data:** Plan

   Why did you select this stage for this task?

Starting data exploration involves planning how to approach data analysis, including understanding data sources and initial data characteristics. This fits the Plan stage.

4. **Data exploration and cleaning:** Analyze **and** Construct

   Why did you select these stages for this task?

Data exploration and cleaning involve analyzing data for patterns and inconsistencies and constructing the cleaned dataset. This involves both analyzing existing data and constructing a usable dataset.

5. **Establish structure for project workflow (PACE):** Plan

Why did you select this stage for this task?

Establishing the workflow involves planning how tasks will be executed and managed throughout the project.

6. **Communicate final insights with stakeholders:** Execute

Why did you select this stage for this task?

Communicating final insights involves presenting results and recommendations to stakeholders, which is part of the Execute stage

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

Computing descriptive statistics involves analyzing the data to summarize its main features, which falls under the Analyze stage.

8. **Visualization building:** Construct **and** Execute

Why did you select these stages for this task?

Building visualizations involves constructing visual representations of data and then using them to execute the presentation of findings.

9. **Write a project proposal:** Plan

Why did you select this stage for this task?

Writing a project proposal involves planning the project structure, goals, and milestones, which fits the Plan stage

10. **Build a regression model:** Construct **and** Execute

Why did you select this stage for this task?

Building a regression model involves constructing the model and executing its use in predictions and evaluations

11. **Compile summary information about the data:** Analyze

Why did you select this stage for this task?

Compiling summary information involves analyzing data to produce a comprehensive overview, fitting the Analyze stage.

12. **Build machine learning model:** Construct

Why did you select this stage for this task?

Building a machine learning model involves constructing the model and executing it in real-world scenarios or testing environments.